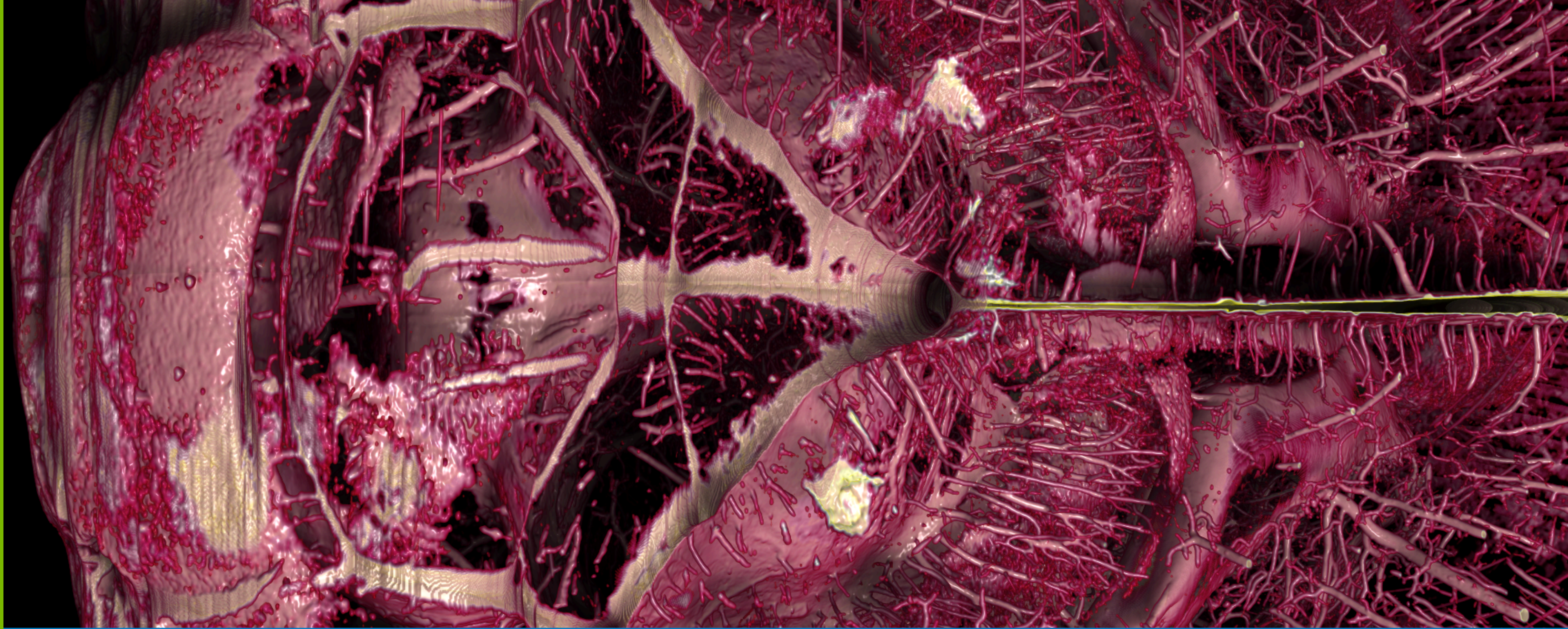
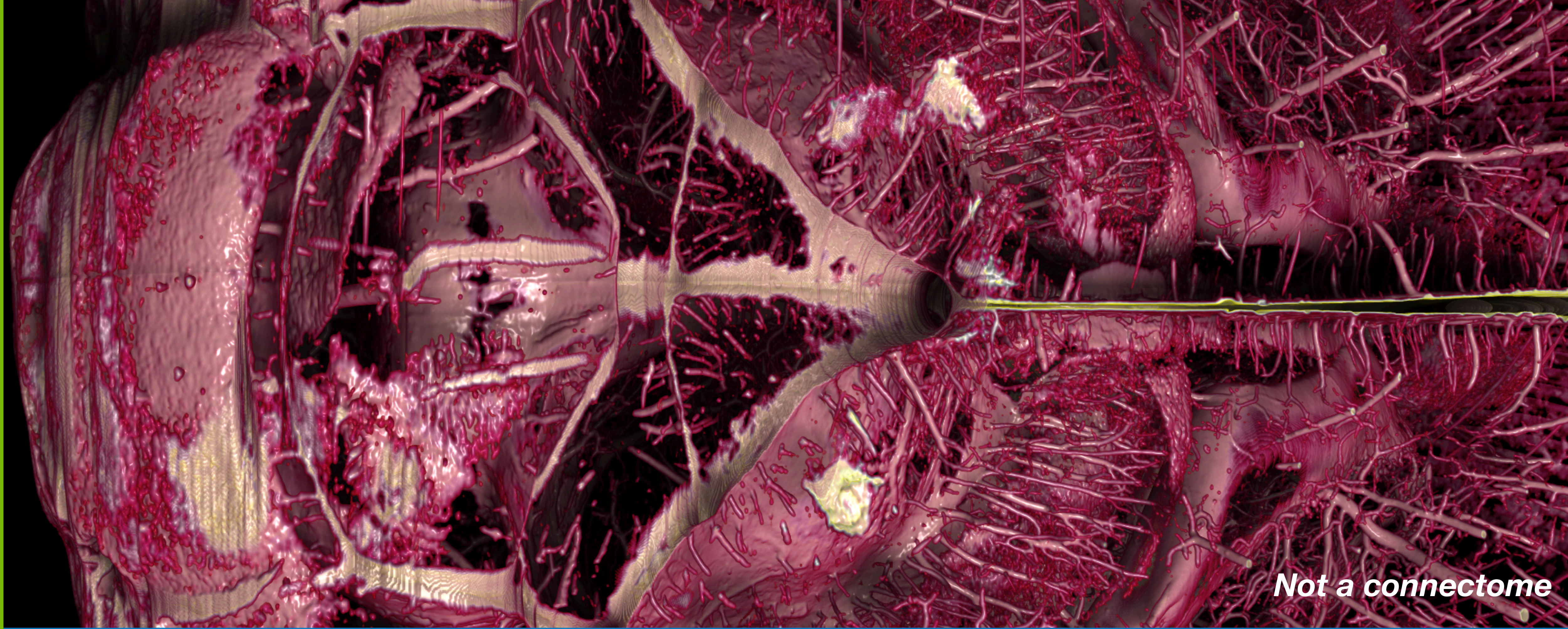


CONNECTOMICS : AURORA EARLY SCIENCE PROJECT

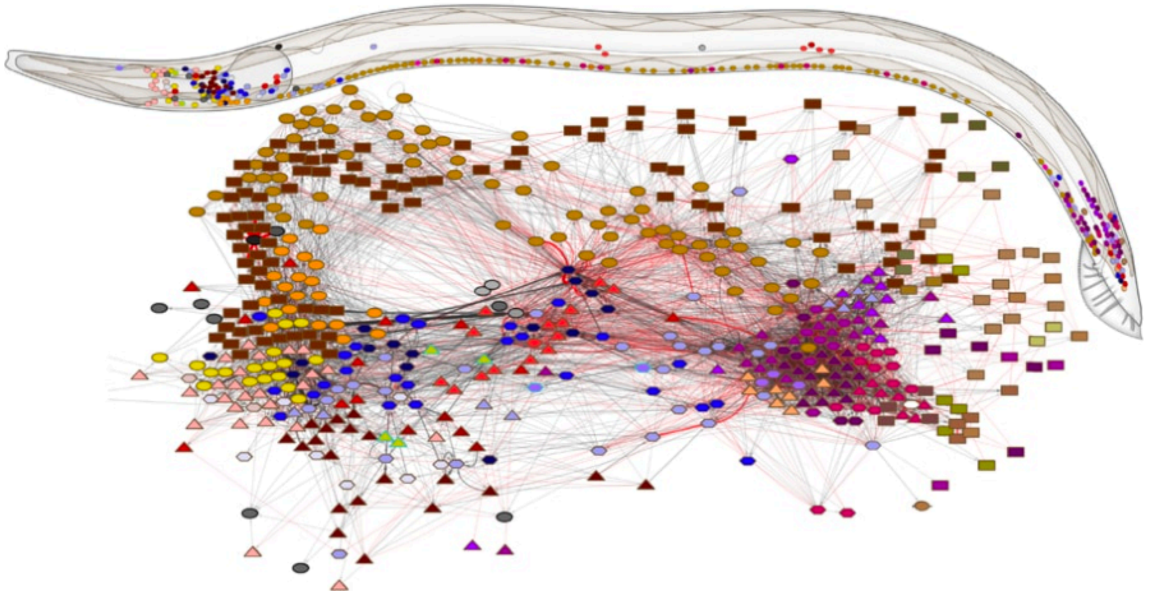
THOMAS URAM, ALCF
TURAM@ANL.GOV



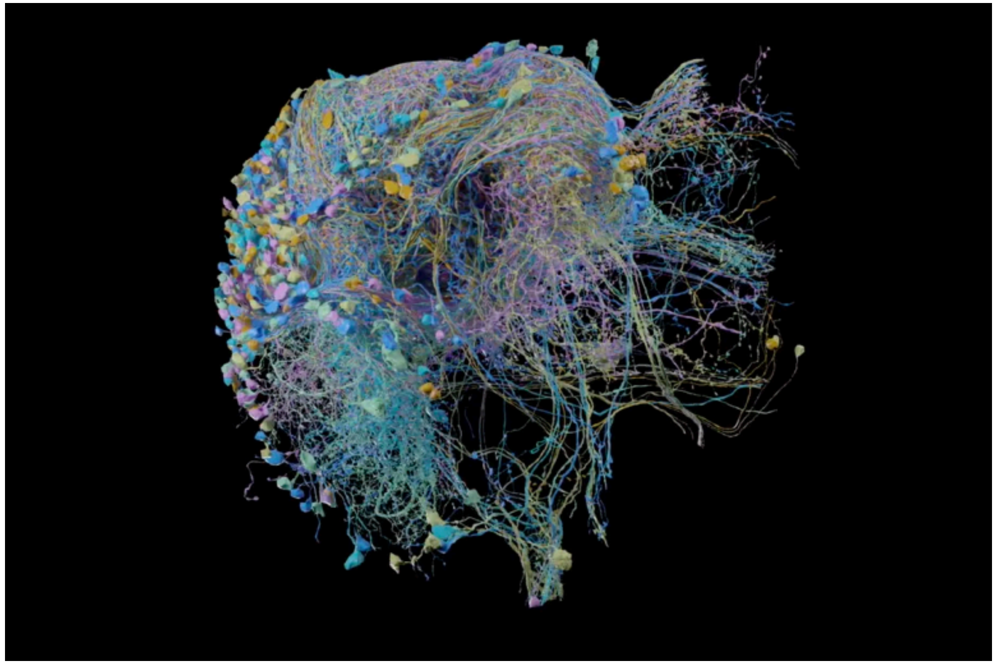
CONNECTOMICS : AURORA EARLY SCIENCE PROJECT

A **connectome** is a comprehensive map of neural connections in the brain; in other words, its "wiring diagram".

CONNECTOMICS SO FAR



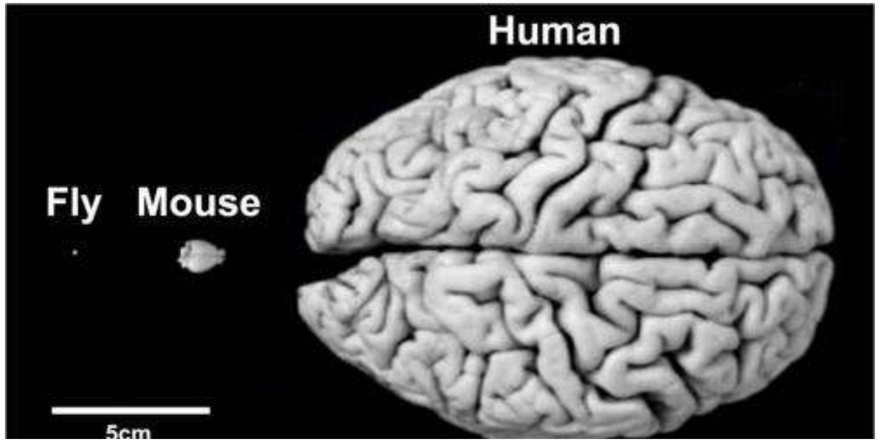
Cook, S.J., Jarrell, T.A., Brittin, C.A. *et al.* Whole-animal connectomes of both *Caenorhabditis elegans* sexes. *Nature* **571,** 63–71 (2019)



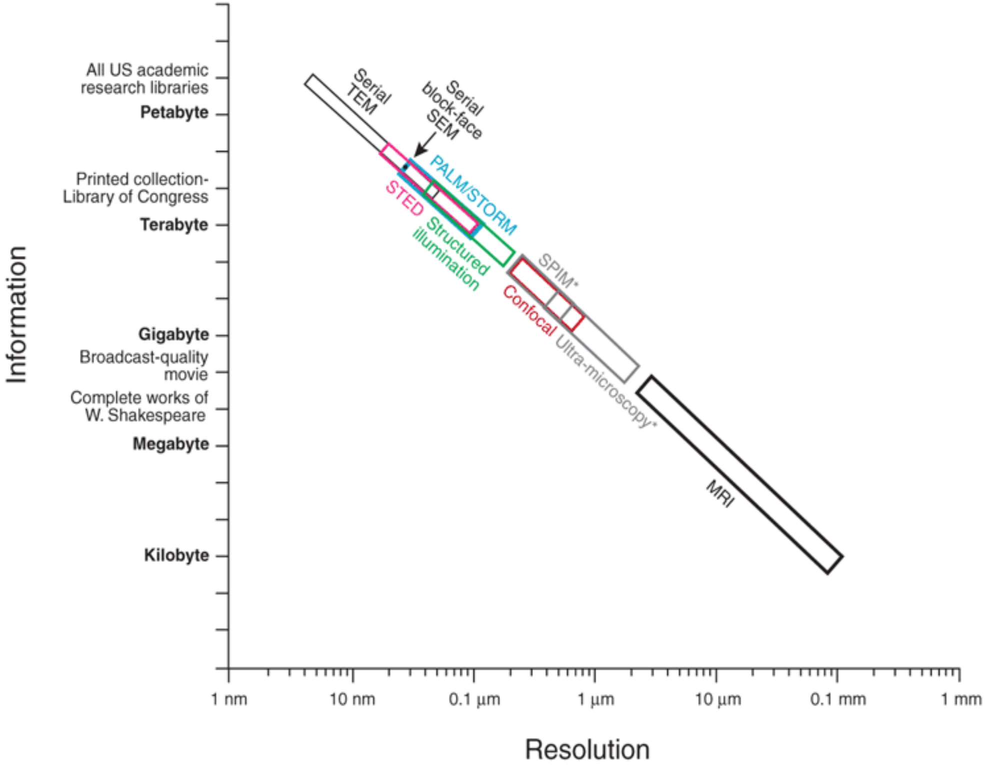
A Connectome of the Adult Drosophila Central Brain

C. Shan Xu, Michal Januszewski, Zhiyuan Lu, Shin-ya Takemura, Kenneth J. Hayworth, Gary Huang, Kazunori Shinomiya, Jeremy Maitin-Shepard, David Ackerman, Stuart Berg, Tim Blakely, John Bogovic, Jody Clements, Tom Dolafi, Philip Hubbard, Dagmar Kainmueller, William Katz, Takashi Kawase, Khaled A. Khairy, Laramie Leavitt, Peter H. Li, Larry Lindsey, Nicole Neubarth, Donald J. Olbris, Hideo Otsuna, Eric T. Troutman, Lowell Umayam, Ting Zhao, Masayoshi Ito, Jens Goldammer, Tanya Wolff, Robert Svirskas, Philipp Schlegel, Erika R. Neace, Christopher J. Knecht Jr., Chelsea X. Alvarado, Dennis A. Bailey, Samantha Ballinger, Jolanta A Borycz, Brandon S. Canino, Natasha Cheatham, Michael Cook, Marisa Dreher, Octave Duclos, Bryon Eubanks, Kelli Fairbanks, Samantha Finley, Nora Forknall, Audrey Francis, Gary Patrick Hopkins, Emily M. Joyce, SungJin Kim, Nicole A. Kirk, Julie Kovalyak, Shirley A. Lauchie, Alanna Lohff, Charli Maldonado, Emily A. Manley, Sari McLin, Caroline Mooney, Miatta Ndama, Omotara Ogundeyi, Nneoma Okeoma, Christopher Ordish, Nicholas Padilla, Christopher Patrick, Tyler Paterson, Elliott E. Phillips, Emily M. Phillips, Neha Rampally, Caitlin Ribeiro, Madelaine K Robertson, Jon Thomson Rymer, Sean M. Ryan, Megan Sammons, Anne K. Scott, Ashley L. Scott, Aya Shinomiya, Claire Smith, Kelsey Smith, Natalie L. Smith, Margaret A. Sobeski, Alia Suleiman, Jackie Swift, Satoko Takemura, Iris Talebi, Dorota Tarnogorska, Emily Tenshaw, Temour Tokhi, John J. Walsh, Tansy Yang, Jane Anne Horne, Feng Li, Ruchi Parekh, Patricia K. Rivlin, Vivek Jayaraman, Kei Ito, Stephan Saalfeld, Reed George, Ian Meinertzhagen, Gerald M. Rubin, Harald F. Hess, Louis K. Scheffer, Viren Jain, Stephen M. Plaza
bioRxiv 2020.01.21.911859; doi: https://doi.org/10.1101/2020.01.21.911859

ENABLING CONNECTOMICS AT EXASCALE TO FACILITATE DISCOVERIES IN NEUROSCIENCE

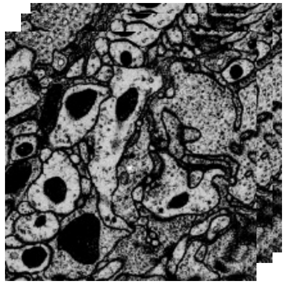


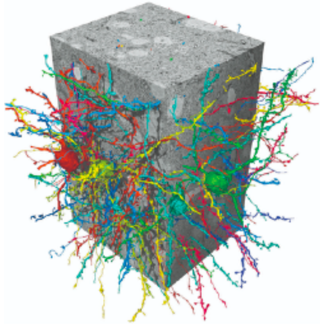
- Why Connectomics?
  - Understand neurodegenerative diseases
  - Understand learning and aging and consciousness
  - Build better, purpose-built neural networks
  - Design neuromorphic computers
- Why Exascale Connectomics?
  - Fruit fly brain: 250K neurons, O(terabytes)
  - Mouse brain: 14M neurons, O(petabytes)
  - Human brain: 100B neurons, O(exabytes)

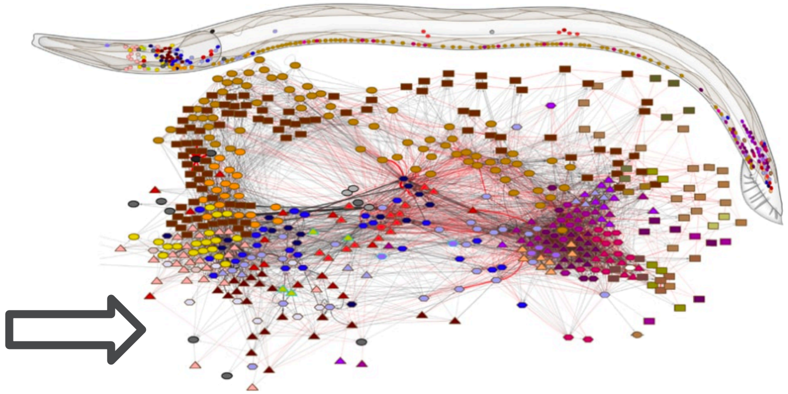


TED talk by Sebastian Seung: https://www.

CONNECTOME CHALLENGES





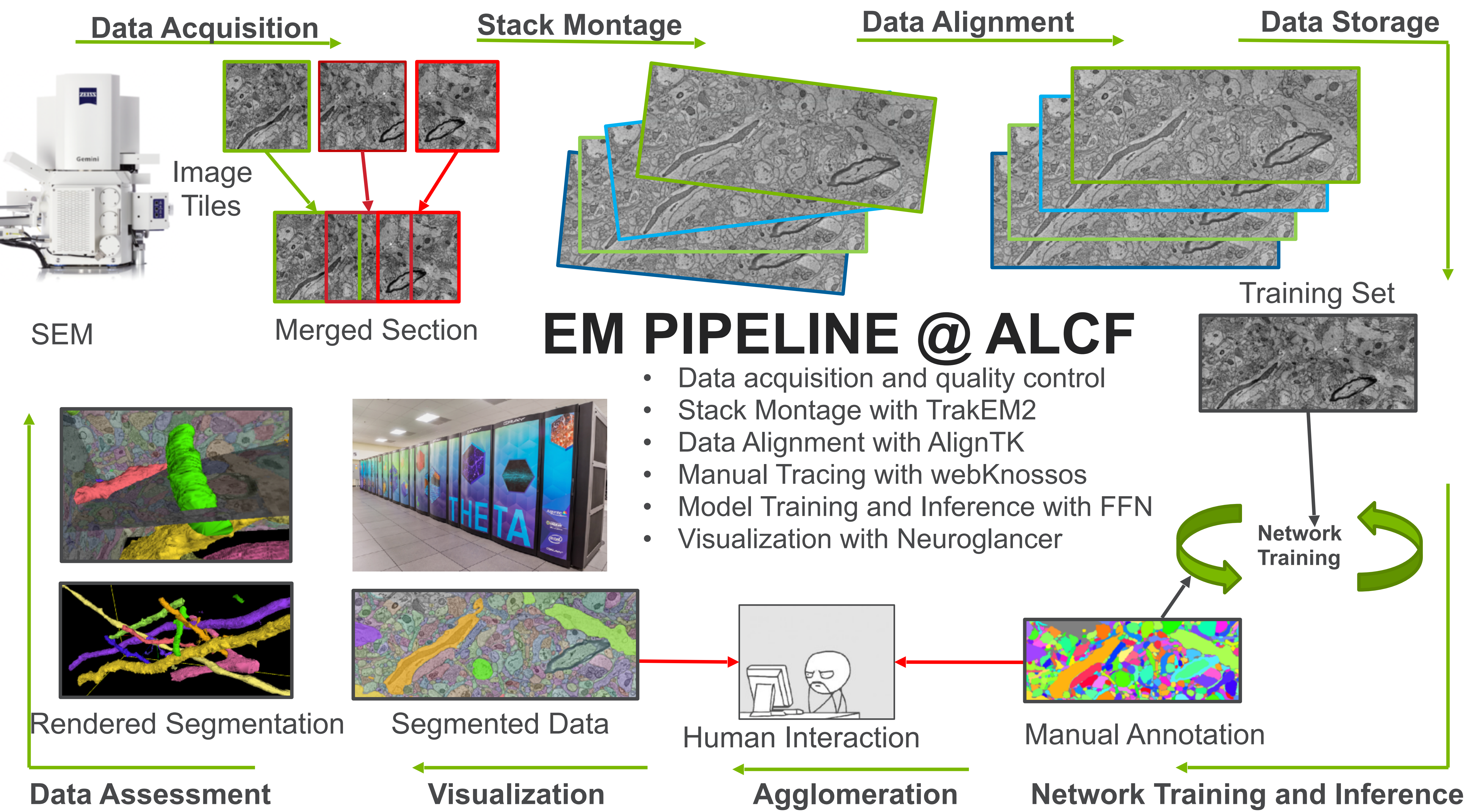


- Big raw data
- Limited training data
  - Significant human effort for ground truth
- Many imperfections or anomalies in data
- Complex neuron shapes in different sizes
- Deep learning model training and large-scale inference

Processing Codes @ANL

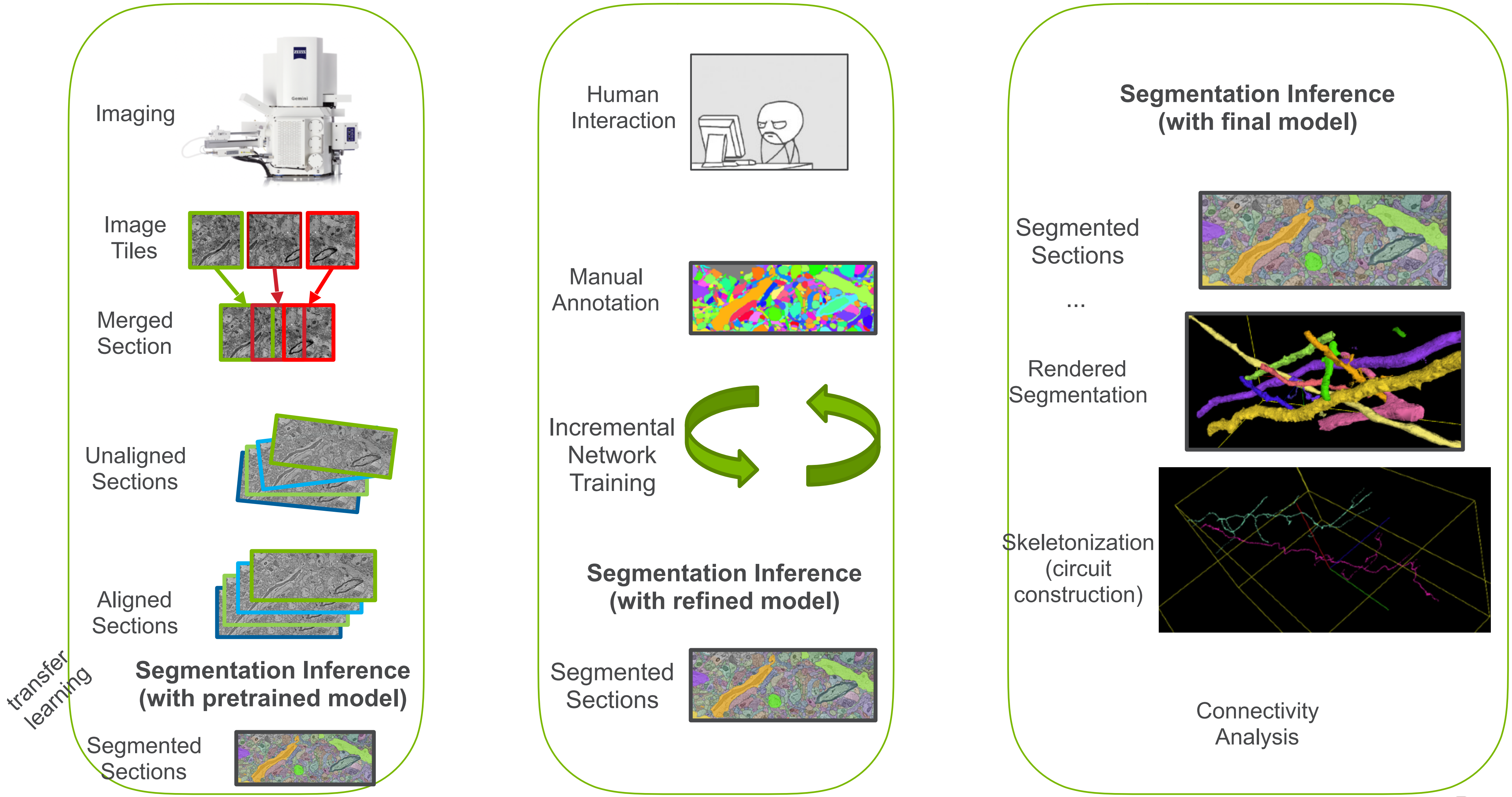
Step	Software	Computing Env
Acquisition	ATLAS	SEM Workstation
Stitching	TrakEM2	Cooley / Theta
Alignment	AlignTK	Theta
Segmentation	**FFN**	**Theta**

Cook, S.J., Jarrell, T.A., Brittin, C.A. *et al.* Whole-animal connectomes of both *Caenorhabditis elegans* sexes. *Nature* **571,** 63–71 (2019)



Data Acquisition
Stack Montage
Data Alignment
Data Storage
Image Tiles
SEM
Merged Section
EM PIPELINE @ ALCF
• Data acquisition and quality control
• Stack Montage with TrakEM2
• Data Alignment with AlignTK
• Manual Tracing with webKnossos
• Model Training and Inference with FFN
• Visualization with Neuroglancer
Training Set
Network Training
THETA
Rendered Segmentation
Segmented Data
Human Interaction
Manual Annotation
Data Assessment
Visualization
Agglomeration
Network Training and Inference

P1: INITIAL SEG   P2: REFINEMENT   P3: FINAL SEG

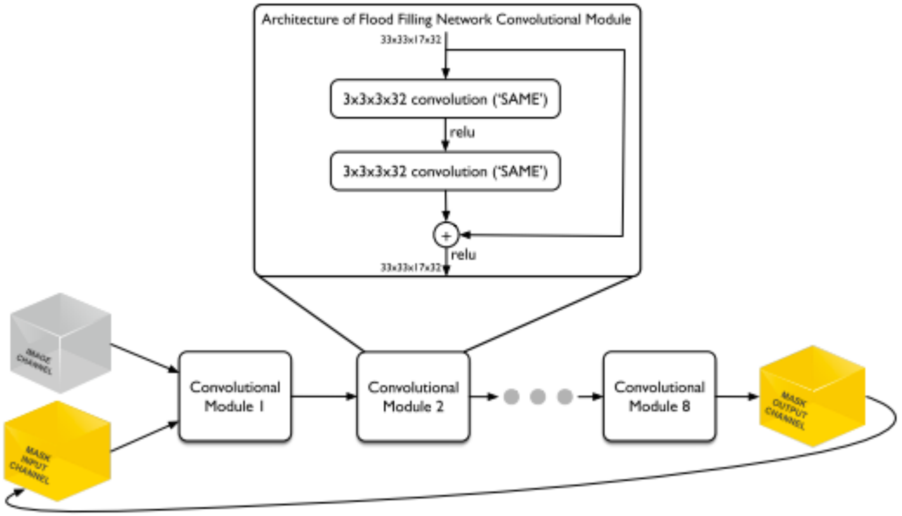


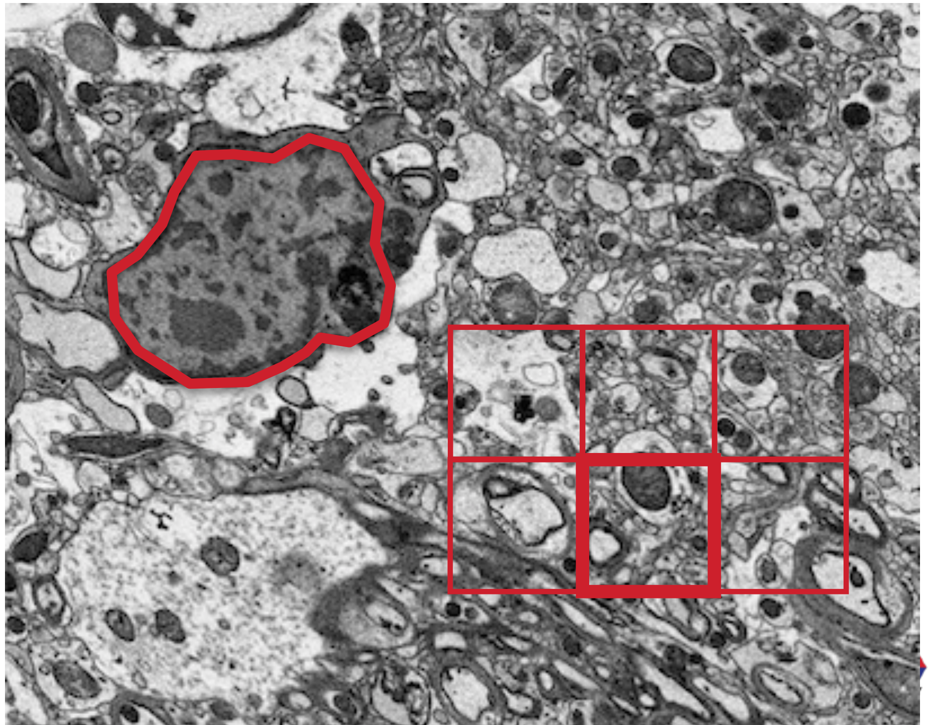
Pipelines built with Balsam (balsam.alcf.anl.gov)

FLOOD FILLING NETWORK (FFN)

TRAINING

- FFN was developed by Google Brain for automated segmentation of structures in EM image data
- Implemented in TensorFlow
- Network consists of a series of 3D convolutional blocks with residual connections
- A 12 layer network has roughly 0.5M trainable parameters
- Network builds on notion of watershed algorithm
    - Find boundaries and fill interior
    - Prefer split errors over merge errors
- Boundary-finding is complicated
    - complex object structure and substructure
    - variation between datasets (limited opportunity for transfer learning, and yet...)
    - variation within datasets (fixing, cutting, staining, imaging)
- Dataset-specific training is required
- Accuracy of FFN is an order of magnitude better than past approaches, but with a higher computational cost.

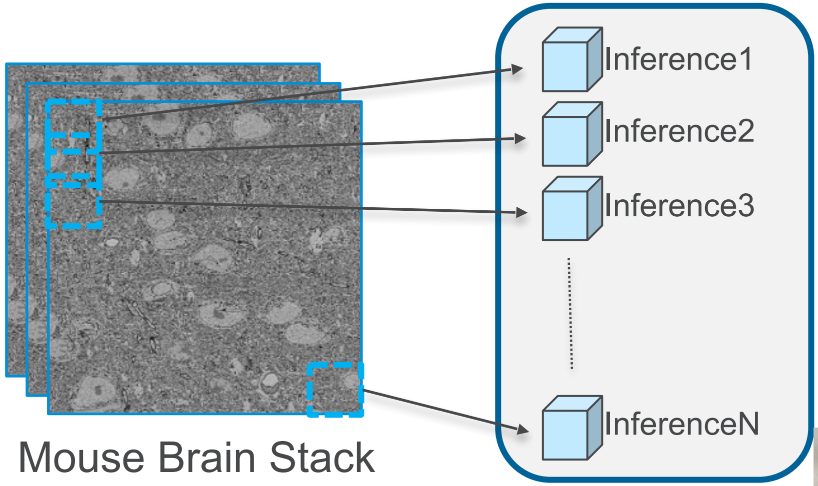




M. Januszewski, J. Kornfeld, P. H. Li, A. Pope, T. Blakely, L. Lindsey, J. Maitin-Shepard, M. Tyka, W. Denk, and V. Jain, “High-precision automated reconstruction of neurons with flood-filling networks,” Nature Methods

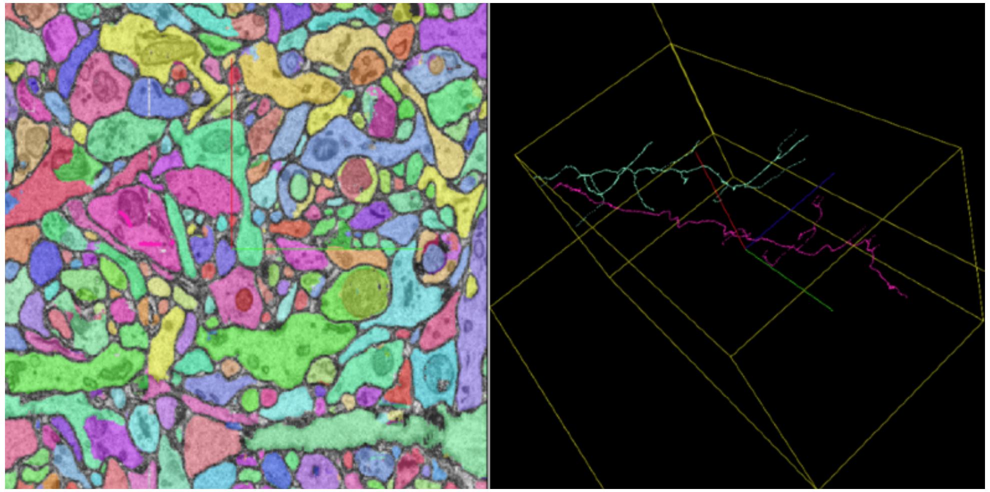
FLOOD FILLING NETWORK (FFN)

INFERENCE



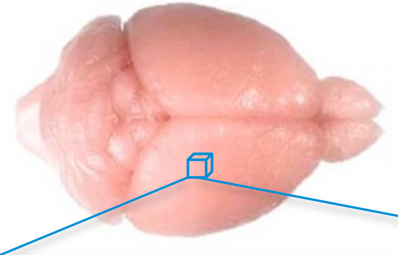
Mouse Brain Stack
21780x15330x1780
~1TB 0.03mm^3
40nm resolution

55,000 inference jobs
256^3 pixels per job
2048 nodes, 16h
(including I/O)



Segmentation and Skeletonization

100x increase in computation
20x increase in I/O speed

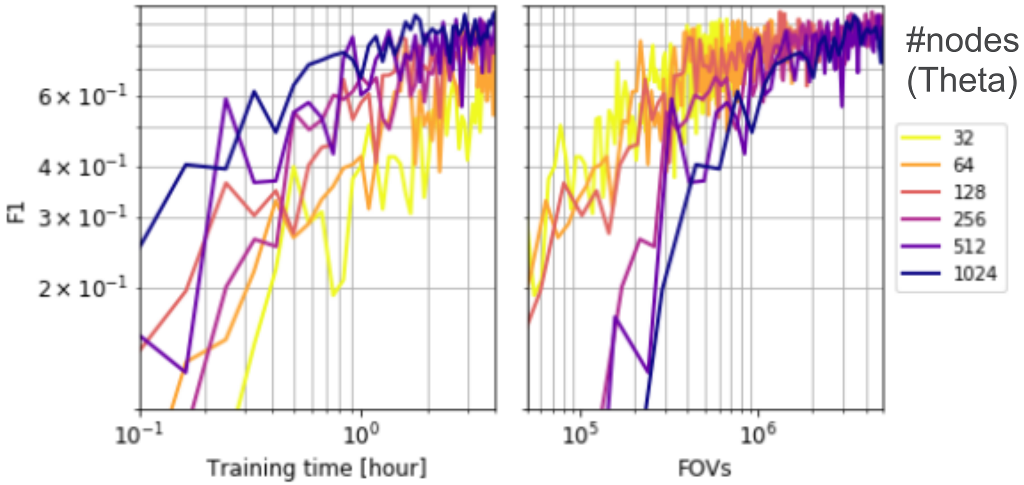


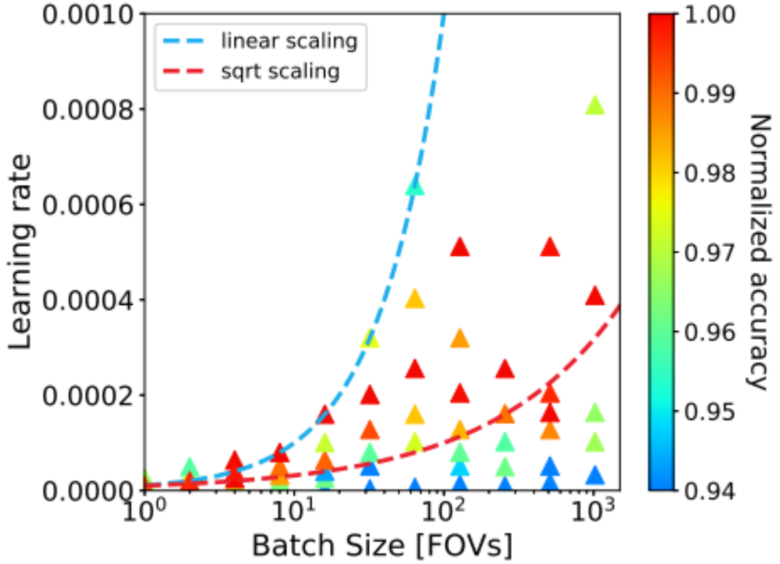
Small fraction of a whole mouse brain

Current processing and I/O challenges will be solved with the upgrade to Aurora. The new system will enable the processing of whole mouse brains at high resolution.

ADVANCES

- Scaling FFN training and inference on ALCF machines
- Initial hyperparameter scans: batch size, learning rate, optimizer, num layers
- Hyperparameter optimization with Deep Hyper (ongoing)
- Evaluation of impact of reduced precision
- Evaluation on future architectures
- Application of learning in other parts of the pipeline (e.g. image alignment)





Wushi Dong, Murat Keceli, Rafael Vescovi, Hanyu Li, Corey Adams, Tom Uram, Venkatram Vishwanath, Bobby Kasthuri, Nicola Ferrier, Peter Littlewood, "Scaling Distributed Training of Flood-Filling Networks on HPC Infrastructure for Brain Mapping", 2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS) at SC19

SUMMARY

- Significant learning and computing challenges remain
- Significant human effort still required
  - How long before learning methods outperform humans on these tasks?
- Obtaining a full mouse connectome will take years of imaging and years of computing
  - Human brain?
- Future will bring advances in microscopes (more pixels!) and methods, and advances in computing (Aurora)
- We are using brain-inspired* neural networks to understand the brain so we can make better neural networks.

** not really*


Argonne
NATIONAL LABORATORY
THE UNIVERSITY OF CHICAGO
1890
PRINCETON
UNIVERSITY
Google